
Academia Open



By Universitas Muhammadiyah Sidoarjo

Table Of Contents

Journal Cover	1
Author[s] Statement.....	3
Editorial Team	4
Article information	5
Check this article update (crossmark)	5
Check this article impact	5
Cite this article.....	5
Title page.....	6
Article Title	6
Author information	6
Abstract	6
Article content	7

Originality Statement

The author[s] declare that this article is their own work and to the best of their knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the published of any other published materials, except where due acknowledgement is made in the article. Any contribution made to the research by others, with whom author[s] have work, is explicitly acknowledged in the article.

Conflict of Interest Statement

The author[s] declare that this article was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright Statement

Copyright © Author(s). This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

Academia Open

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14903

EDITORIAL TEAM

Editor in Chief

Mochammad Tanzil Multazam, Universitas Muhammadiyah Sidoarjo, Indonesia

Managing Editor

Bobur Sobirov, Samarkand Institute of Economics and Service, Uzbekistan

Editors

Fika Megawati, Universitas Muhammadiyah Sidoarjo, Indonesia

Mahardika Darmawan Kusuma Wardana, Universitas Muhammadiyah Sidoarjo, Indonesia

Wiwit Wahyu Wijayanti, Universitas Muhammadiyah Sidoarjo, Indonesia

Farkhod Abdurakhmonov, Silk Road International Tourism University, Uzbekistan

Dr. Hindarto, Universitas Muhammadiyah Sidoarjo, Indonesia

Evi Rinata, Universitas Muhammadiyah Sidoarjo, Indonesia

M Faisal Amir, Universitas Muhammadiyah Sidoarjo, Indonesia

Dr. Hana Catur Wahyuni, Universitas Muhammadiyah Sidoarjo, Indonesia

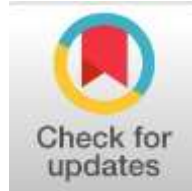
Complete list of editorial team ([link](#))

Complete list of indexing services for this journal ([link](#))

How to submit to this journal ([link](#))

Article information

Check this article update (crossmark)



Check this article impact (*)



Save this article to Mendeley



(*) Time for indexing process is various, depends on indexing database platform

Semantic Analysis of Quiz Based Educational Platforms Using Pre Trained Models: Analisis Semantik Platform Pendidikan Berbasis Kuis Menggunakan Model yang Telah Dilatih Sebelumnya

Muhamammad Zaky Novananda, muhammad.zaky.2205356@students.um.ac.id (*)

Program Studi Teknik Informatika, Universitas Negeri Malang, Indonesia

Triyanna Widiyaningtyas, triyannaw.ft@um.ac.id

Program Studi Teknik Informatika, Universitas Negeri Malang, Indonesia

(*) Corresponding author

Abstract

General Background: Teknologi digital telah merevolusi pendidikan melalui platform edukasi berbasis kuis yang interaktif dan fleksibel. **Specific Background:** Minat remaja terhadap platform ini ditentukan oleh persepsi visual, kemudahan operasional, dan kegunaan fungsional. **Knowledge Gap:** Kurangnya integrasi antara analisis regresi kuantitatif dan eksplorasi semantik berbasis NLP membatasi pemahaman mendalam mengenai pengalaman pengguna. **Aims:** Penelitian ini bertujuan menganalisis faktor penentu minat pengguna serta mengevaluasi model BERT, SBERT, dan DistilBERT dalam memproses opini pengguna secara semantik. **Results:** Analisis regresi ($R^2 = 0,7516$) mengidentifikasi kegunaan fungsional sebagai prediktor utama minat pengguna, sementara BERT menunjukkan performa terbaik dengan skor Cosine Similarity tertinggi (0,7666) dibanding model lainnya. **Novelty:** Penelitian ini mengintegrasikan desain metode campuran untuk mengukur pendorong keterlibatan pengguna sekaligus membuktikan keunggulan arsitektur BERT dalam menginterpretasi respons pengguna di lingkungan pendidikan. **Implications:** Temuan ini menyediakan kerangka kerja berbasis data bagi pengembang untuk memprioritaskan kegunaan fungsional dan memanfaatkan analisis semantik BERT dalam optimalisasi platform edukasi berkelanjutan.

Highlights

- ♦ Functional utility emerged as the most significant predictor of platform engagement, surpassing visual design and operational ease.
- ♦ BERT demonstrated superior performance in capturing semantic similarity within user feedback compared to DistilBERT and SBERT.
- ♦ Mixed-methods integration effectively provides a comprehensive evaluation of user perception for targeted educational system improvements.

Keywords

Educational Website; Interactive Quiz; BERT; DistilBERT; SBERT; NLP

Published date: 2026-07-17

PENDAHULUAN

Perkembangan teknologi informasi yang pesat saat ini mulai membawa perubahan besar dalam memperoleh, mengakses dan mengolah informasi. Salah satu bidang yang paling berpengaruh dari dampak perkembangan teknologi yaitu pendidikan. Perkembangan digital dalam dunia pendidikan banyak memberikan media pembelajaran yang interaktif dan fleksibel. Salah satunya adalah dengan platform edukasi berbasis web. Pembelajaran seperti ini memungkinkan proses belajar secara lebih efektif tanpa dibatasi ruang dan waktu [1].

Penggunaan media interaktif berbasis web dapat meningkatkan minat dan motivasi belajar siswa. Selain itu, web-based assessment mampu menumbuhkan ketertarikan siswa terhadap proses pembelajaran karena memberikan pengalaman belajar yang lebih fleksibel dan menarik dibandingkan metode konvensional. Media pembelajaran interaktif juga berperan dalam menciptakan suasana belajar yang menyenangkan, sehingga dapat mendukung keterlibatan aktif siswa selama proses pembelajaran. Pembelajaran yang disajikan secara digital memungkinkan siswa berinteraksi lebih aktif dengan materi. Hal ini diperkuat oleh temuan penelitian, bahwa media pembelajaran interaktif berbasis hypertext secara signifikan dapat meningkatkan minat baca siswa dalam konteks pembelajaran daring [2].

Di sisi lain, era digital juga menghadirkan tantangan tersendiri. Penggunaan gadget secara berlebihan oleh siswa kerap berdampak negatif terhadap proses belajar dan keterampilan sosial mereka. Ketergantungan terhadap gadget dapat menyebabkan siswa cenderung mengabaikan proses belajar dan lebih memilih bergantung pada teknologi. Data menunjukkan bahwa penggunaan gadget pada anak meningkat dari 38% pada tahun 2011 menjadi 72% pada tahun 2013, seiring dengan kemajuan teknologi seperti fitur touch screen yang memudahkan penggunaan [3]. Meski gadget memiliki manfaat seperti mendukung pembelajaran daring, menyediakan akses ke kuis dan tutorial online, serta melatih keterampilan kognitif dan motorik. Dampak negatifnya pun tidak bisa diabaikan, diantaranya adalah meningkatnya sifat individualis, terhambatnya perkembangan emosional, serta menurunnya kemampuan komunikasi dan sosialisasi [4].

Saat ini, banyak remaja dan pengguna umum cenderung lebih banyak menghabiskan waktu dengan perangkat digital dibandingkan membaca buku atau mempelajari materi konvensional dari sekolah. Kondisi ini menghadirkan tantangan sekaligus peluang bagi dunia pendidikan untuk menciptakan media pembelajaran yang informatif dan juga menarik. Upaya ini diharapkan mampu meningkatkan minat belajar siswa di tengah perkembangan teknologi. Salah satu pendekatan yang potensial adalah penggunaan website edukatif berbasis soal interaktif. Media pembelajaran sederhana berbasis web, seperti Google Sites, terbukti efektif dalam meningkatkan minat belajar siswa pada mata pelajaran tertentu [5].

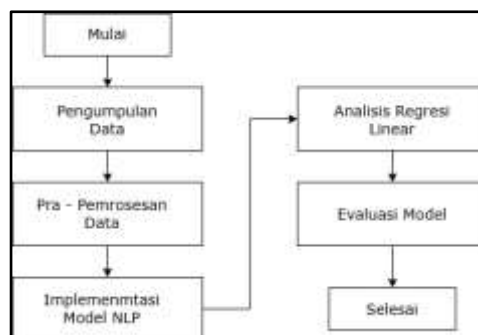
Penulis bersama dengan tim pengembang telah mengembangkan sebuah prototype website edukatif yang berisi kumpulan soal interaktif untuk melatih kemampuan pengguna. Website ini dikembangkan dengan tampilan menarik, fitur penilaian otomatis, dan berorientasi pada pengalaman pengguna. Melalui pendekatan kuantitatif berbasis kuesioner, penelitian ini mengkaji minat dan respons pengguna sebagai dasar pengembangan sistem pembelajaran digital yang lebih efektif, adaptif, dan sesuai kebutuhan, sekaligus mendukung pengembangan media edukasi berbasis teknologi di era digital.

METODE

A. Jenis dan Pendekatan Penelitian

Penelitian ini mengimplementasikan pendekatan mixed-methods eksplanatoris sekuensial yang mengintegrasikan analisis kuantitatif melalui regresi linear berganda dan analisis kualitatif melalui eksplorasi semantik teks menggunakan tiga model pre-trained NLP, yaitu BERT, SBERT, dan DistilBERT. Penelitian ini dilakukan di lingkungan Universitas Negeri Malang pada periode April hingga Mei 2026.

Gambar 1 Flowchart Alur Penelitian



Variabel independen dalam penelitian ini meliputi tampilan visual (X_1), kegunaan fungsional (X_2), dan kemudahan operasional (X_3), sedangkan variabel dependen adalah minat pengguna (Y). Seluruh variabel diukur menggunakan skala Likert 1–5 melalui kuesioner daring yang disebar kepada 74 responden mahasiswa Universitas Negeri Malang. Instrumen terdiri atas pertanyaan tertutup untuk analisis regresi dan pertanyaan terbuka untuk analisis semantik NLP.

Ketiga model NLP digunakan untuk menganalisis kesamaan semantik jawaban terbuka responden melalui pendekatan Semantic Textual Similarity (STS) dengan metrik Cosine Similarity. BERT digunakan dengan konfigurasi bert-base-multilingual-cased, SBERT dengan arsitektur siamese network, dan DistilBERT sebagai versi ringkas BERT dengan knowledge distillation. Analisis regresi linear berganda dilakukan untuk mengukur pengaruh ketiga variabel independen

terhadap minat pengguna, sedangkan evaluasi model menggunakan MSE, RMSE, dan koefisien determinasi (R^2) untuk mengetahui kemampuan model dalam menjelaskan variasi data [6].

HASIL DAN PEMBAHASAN

A. Hasil Pengumpulan Data

Pengumpulan data dilakukan melalui penyebaran kuesioner daring menggunakan Google Form kepada remaja terutama mahasiswa selama periode penelitian. Data diperoleh sebanyak 74 responden yang memenuhi kriteria penelitian. Seluruh responden berpengalaman menggunakan platform kuis online. Karakteristik responden penelitian dapat dilihat pada Tabel 1

Tabel 1 Karakteristik Responden Penelitian

Karakteristik	Kategori	Jumlah	Presentase (%)
Jenis Kelamin	Perempuan	39	52,7%
	Laki – laki	35	47,3%
Angkatan	2022	46	62,2%
	2023	9	12,2%
	2024	11	14,9%
	2025	8	10,8%
Usia	Rata-rata	21,28 tahun	-
	Rentang	18-25 tahun	-
Penggunaan Kuis Online	Pernah	74	100%

Berdasarkan tabel 1 menunjukkan bahwa, responden didominasi oleh mahasiswa perempuan sebanyak 39 orang (52,7%), sedangkan mahasiswa laki-laki sebanyak 35 orang (47,3%). Mayoritas responden berasal dari Angkatan 2022 sebanyak 46 orang (62,2%), diikuti Angkatan 2024 sebanyak 11 orang (14,9%), Angkatan 2023 sebanyak 9 orang (12,2%), dan Angkatan 2025 sebanyak 8 orang (10,8%). Rata-rata usia responden 21,28 tahun (18-25 tahun), yang mencerminkan karakteristik digital native dengan tingkat kedekatan tinggi terhadap platform digital. Hal ini menunjukkan bahwa media sosial telah menjadi bagian dari gaya hidup sekaligus sarana berekspresi [7].

B. Statistik Deskriptif Skor Likert

Analisis statistik deskriptif dilakukan terhadap data hasil kuesioner tertutup yang menggunakan skala Likert dengan nilai 1 hingga 5, untuk nilai 1 berarti sangat Tidak Setuju dan nilai 5 berarti Sangat Setuju. Kuesioner mencakup empat aspek: Tampilan Visual (5 butir), Kemudahan Operasional (6 butir), Kegunaan Fungsional (5 butir), dan Minat Pengguna (2 butir). Hasilnya disajikan pada Tabel 2.

Tabel 2 Statistik Dekskriptif Skor Likert per Aspek

Aspek	Rata-rata	Std Dev	Min	Max	Kategori
Tampilan Visual	3,98	0,80	1,20	5,00	Sedang
Kemudahan Operasional	3,88	0,79	1,67	5,00	Sedang
Kegunaan Fungsional	4,00	0,78	1,00	5,00	Sedang
Minat Pengguna	3,86	0,97	1,00	5,00	Sedang

Berdasarkan Tabel 2, menunjukkan seluruh aspek yang dievaluasi memperoleh rata-rata skor pada kategori sedang dengan rentang nilai 3,86–4,00. Hasil ini menunjukkan bahwa responden memberikan penilaian yang positif terhadap website edukasi berbasis kuis yang dikembangkan. Secara umum, platform dinilai mampu memenuhi kebutuhan pengguna baik dari segi tampilan, kemudahan penggunaan, maupun manfaat yang diberikan dalam mendukung proses pembelajaran.

Aspek kegunaan fungsional memperoleh rata-rata tertinggi ($M = 4,00$), menunjukkan bahwa platform bermanfaat dalam mendukung pembelajaran dan menghadirkan pengalaman belajar yang lebih interaktif. Aspek tampilan visual memperoleh rata-rata 3,98, yang mengindikasikan bahwa desain antarmuka telah memberikan kenyamanan bagi pengguna.

Aspek kemudahan operasional ($M = 3,88$) dan minat pengguna ($M = 3,86$) juga memperoleh penilaian positif. Hasil ini menunjukkan bahwa platform mudah digunakan dan berpotensi digunakan Kembali, meskipun variasi pada aspek minat mencerminkan perbedaan tingkat ketertarikan antarresponden terhadap pemanfaatannya sebagai media pembelajaran.

Secara keseluruhan, hasil analisis deskriptif menunjukkan bahwa website edukasi berbasis kuis yang dikembangkan telah diterima dengan baik oleh pengguna. Temuan ini mengindikasikan bahwa platform memiliki potensi untuk digunakan sebagai media pembelajaran yang mendukung proses belajar secara lebih efektif, menarik, dan mudah diakses.

C. Analisis Regresi Linear Berganda

Analisis regresi linear berganda digunakan untuk menjawab masalah penelitian, yaitu digunakan untuk mengukur pengaruh tampilan visual (X_1), kemudahan operasional (X_2), dan kegunaan fungsional (X_3) terhadap minat pengguna (Y). Seluruh proses perhitungan dilakukan menggunakan pustaka scikit-learn pada Bahasa pemrograman. Hasil analisis regresi linear berganda dapat dilihat pada pada Tabel 3.

ISSN 2714-7444 (online), <https://acopen.umsida.ac.id>, published by Universitas Muhammadiyah Sidoarjo

Tabel 3 Hasil Analisis Regresi Linear Berganda

Variabel	Koefisien (β)	Keterangan
Konstanta (α)	-0,3588	Nilai dasar model regresi
Tampilan Visual (X1)	0,0363	Pengaruh positif terhadap Minat
Kemudahan Operasional (X2)	-0,0913	Tertekan oleh dominasi Kegunaan
Kegunaan Operasional (X3)	1,1091	Pengaruh paling dominan
R ² (Koefisien Determinasi)	0,7516	Model menjelaskan 75,2% variansi minat
MSE	0,2288	Rata-rata kuadrat kesalahan prediksi
RMSE	0,4783	Akar rata-rata kuadrat kesalahan

Berdasarkan tabel 3, terlihat bahwa ketiga variabel yang diteliti, yaitu Tampilan Visual (X1), Kemudahan Operasional (X2) dan kegunaan Operasional (X3) memiliki kontribusi yang berbeda terhadap Minat Pengguna (Y). Hasil ini menunjukkan bahwa minat pengguna terhadap website edukasi berbasis kuis tidak hanya dipengaruhi oleh satu faktor, tetapi kombinasi dari pengalaman pengguna saat berinteraksi dengan platform. Hasil regresi di atas menunjukkan bahwa aspek Kegunaan Operasional memiliki pengaruh paling dominan dibandingkan aspek lainnya. Hal tersebut dipengaruhi oleh mayoritas pengguna yang memberikan penilaian di rata – rata angka 4, yang merupakan rata – rata tertinggi dari aspek yang lain. Hasil tersebut mengindikasikan bahwa pengguna lebih mempertimbangkan manfaat yang diperoleh dari platform dibandingkan aspek tampilan maupun kemudahan penggunaan. Nilai koefisien determinasi (R^2) sebesar 0,7516 yang tergolong cukup tinggi menunjukkan bahwa model memiliki kemampuan yang baik dalam menjelaskan faktor-faktor yang memengaruhi minat pengguna. Nilai Mean Squared Error (MSE) yang relatif kecil sebesar 0,2288 mengindikasikan bahwa selisih antara hasil prediksi model dan kondisi sebenarnya tidak terlalu besar. Nilai Root Mean Squared Error (RMSE) sebesar 0,4783 menunjukkan bahwa rata-rata kesalahan prediksi model berada pada kisaran 0,48 poin tergolong rendah dari skala penilaian 1-5, sehingga model regresi yang dihasilkan memiliki tingkat akurasi prediksi yang cukup baik dan mampu menggambarkan hubungan antara karakteristik platform edukasi dengan minat pengguna secara memadai.

Variabel Kegunaan Fungsional (X3) memiliki koefisien regresi terbesar yaitu $\beta_3 = 1,109$, yang menunjukkan bahwa setiap peningkatan satu poin pada skor kegunaan akan meningkatkan minat pengguna sebesar 1,109 poin, dengan asumsi variabel lain dianggap konstan. Dominasi variabel Kegunaan ini diperkuat oleh hasil analisis korelasi yang menunjukkan nilai korelasi sangat tinggi antara Kegunaan dan Minat ($r = 0,864$), sebagaimana ditampilkan pada Tabel 4.8, artinya jika variabel independen lain nilainya tetap dan persepsi kegunaan mengalami kenaikan atau yang jumlahnya tinggi maka koefisien bernilai positif artinya terjadi hubungan searah antara variabel persepsi kegunaan dan minat penggunaan, semakin baik persepsi Kegunaan diimbangi dengan tingginya Minat pengguna (Alexander & Suhendry, 2024).[8]

Tabel 4 Matriks Korelasi Antar Variabel Penelitian

Variabel	Tampilan	Kemudahan	Kegunaan	Minat
Tampilan	1,000	0,344	0,531	0,477
Kemudahan	0,344	1,000	0,555	0,430
Kegunaan	0,531	0,555	1,000	0,864
Minat	0,477	0,430	0,864	1,000

Tampilan visual ($\beta_1 = 0,036$) berpengaruh positif meski kecil, karena desain yang menarik berkontribusi pada kesan pertama namun bukan penentu utama penggunaan berulang (Amesha et al., 2025). Secara keseluruhan, pengembangan platform sebaiknya diprioritaskan pada peningkatan kegunaan fungsional seperti relevansi konten, kedalaman materi, dan manfaat pembelajaran yang dirasakan langsung sebagai kunci utama membangun dan mempertahankan minat pengguna. Variabel Kemudahan Operasional (X2) menunjukkan koefisien negatif sebesar $\beta_2 = -0,091$. Ketika variabel Kegunaan memiliki korelasi yang sangat tinggi terhadap Minat ($r = 0,864$) dan korelasi yang cukup signifikan terhadap Kemudahan ($r = 0,555$), kondisi menunjukkan bahwa dalam konteks dataset penelitian ini, pengaruh kemudahan terhadap minat sudah tercakup secara implisit di dalam variabel Kegunaan yang lebih dominan [9].

Hasil regresi menunjukkan bahwa kegunaan fungsional merupakan faktor dominan yang memengaruhi minat pengguna, sehingga pengembangan platform perlu memprioritaskan peningkatan kualitas konten, materi, dan manfaat pembelajaran.

D. Implementasi Model NLP

Analisis respons uraian terbuka dilakukan menggunakan tiga model NLP pre-trained dari HuggingFace, yaitu BERT, SBERT, dan DistilBERT. Ketiganya dipilih karena mendukung pemrosesan Bahasa Indonesia melalui kemampuan multilingual dan merepresentasikan tiga pendekatan berbeda dalam menghasilkan sentence embeddings [10]. Konfigurasi teknis masing-masing model disajikan pada Tabel 5.

Tabel 5 Konfigurasi Model Pre-trained NLP

Model	Nama Model	Arsitektur	Pendekatan
BERT	HuggingFace Bert-base-multilingual-cased	Bidirectional Transformer Encoder	Sentence-Embedding via Mean Polling
SBERT	Paraphrase-multilingual-miniLM-LV12-v2	Siamese BERT Network	Optimized Sentence Embedding
DistilBERT	Paraphrase-multilingual-mpnet-v2	DistilBERT XLM-RoBERTa	Knowledge Distillation Embedding

1. Pendekatan Semantic Textual Similarity

Pendekatan Semantic Textual Similarity (STS) dipilih karena data berupa kalimat deskriptif, bukan teks berpolaritas emosi. STS mengubah kalimat menjadi sentence embedding dan mengukur kemiripan makna menggunakan Cosine Similarity (0-1), dengan nilai yang mendekati 1 menunjukkan kemiripan semantik yang tinggi. Cosine Similarity (A, B) = $(A \cdot B) / (||A|| \times ||B||)$

Nilai Cosine Similarity berkisar antara 0 hingga 1. Nilai yang mendekati 1 mengindikasikan bahwa kedua teks memiliki makna yang sangat mirip atau hampir identik, sedangkan nilai yang mendekati 0 mengindikasikan bahwa kedua teks memiliki makna yang sangat berbeda satu sama lain [11]. Dalam konteks penelitian ini, nilai Cosine Similarity yang tinggi antar jawaban responden menunjukkan adanya konsensus atau kesamaan persepsi di antara pengguna terhadap aspek tertentu dari platform.

2. Proses Implementasi

Proses implementasi diawali dengan pemuatan model pre-trained beserta tokenizer bawaan masing-masing dari repositori HuggingFace. Dari 73 teks valid, dipilih 30 pasangan secara acak (random seed 42) dari 2.628 pasangan yang mungkin untuk efisiensi komputasi. Nilai Cosine Similarity dihitung untuk setiap pasangan menggunakan ketiga model secara identik agar perbandingan performa berlangsung adil [12].

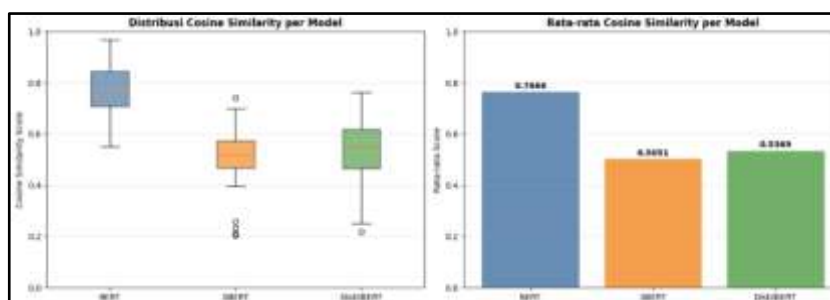
E. Perbandingan Performa Model

Hasil perbandingan performa ketiga model NLP dalam menghasilkan representasi semantik dari teks jawaban responden disajikan secara lengkap pada Tabel 4.10. Perbandingan dilakukan menggunakan empat metrik statistik yaitu rata-rata Cosine Similarity Score, standar deviasi, skor tertinggi, dan skor terendah pasangan teks yang dianalisis. Keempat metrik tersebut memberikan gambaran komprehensif atas performa setiap model, mencakup nilai rata-rata, konsistensi, serta rentang kemampuan diskriminasi semantiknya [13].

Tabel 6 Perbandingan Cosine Similarity Score Model NLP

Model	Rata-rata Score	Std Deviasi	Score Tertinggi	Score Terendah	Rank
BERT	0,766	0,1042	0,9681	0,5493	1
DistilBERT	0,5369	0,1319	0,7636	0,2167	2
SBERT	0,5051	0,1375	0,7419	0,2040	3

Gambar 3 Grafik Distribusi dan Rata-rata Cosine Similarity



Berdasarkan tabel 6, BERT menunjukkan performa terbaik dengan rata-rata cosine similarity sebesar 0,7666 dan standar deviasi terendah (0,1042), yang mengindikasikan tingkat akurasi dan konsistensi yang lebih baik dibandingkan model lainnya. BERT memperoleh skor tertinggi (0,9681), menunjukkan kemampuan terbaik dalam menangkap kesamaan makna, sejalan dengan Mudding (2024).

Sebaliknya, DistilBERT (0,5369) dan SBERT (0,5051) juga mampu merepresentasikan kesamaan semantik tetapi performanya masih di bawah BERT. Hasil tersebut konsisten dengan penelitian Putri et al. (2023) dan Sungkono et al. (2023) yang menunjukkan bahwa efektivitas model sangat dipengaruhi oleh kemampuan arsitektur dalam memahami konteks serta karakteristik data yang digunakan [14][15]. Dengan demikian, BERT dipilih sebagai model terbaik pada penelitian ini karena menghasilkan tingkat kesamaan semantik yang paling tinggi dan konsisten.

Model DistilBERT menempati posisi kedua dengan rata-rata skor 0,5369 dan standar deviasi 0,1319. Meskipun performanya berada di bawah BERT, hasil tersebut menunjukkan bahwa DistilBERT masih mampu mempertahankan kualitas representasi semantik yang cukup baik. Hal ini sejalan dengan penelitian Putri et al. (2023), bahwa DistilBERT mampu melakukan proses pelatihan dan prediksi lebih cepat serta lebih efisien dalam penggunaan memori dengan penuruna akurasi yang relative kecil dibandingkan BERT, sehingga DistilBERT dapat menjadi alternatif yang efektif ketika efisiensi komputasi menjadi pertimbangan utama.

1. Performa Model BERT

Model BERT menunjukkan performa terbaik dengan rata-rata cosine similarity 0,7666 dan standar deviasi terendah (0,1042), yang mencerminkan kemampuan akurat dan konsisten dalam mengukur kesamaan makna. Hasil ini menunjukkan bahwa representasi semantik yang dihasilkan BERT lebih stabil dibandingkan model lainnya. Selain itu, skor tertinggi sebesar 0,9681 menunjukkan adanya pasangan teks yang hampir identik secara semantik, sedangkan skor terendah sebesar 0,5493 mengindikasikan bahwa bahkan pada pasangan teks yang paling berbeda, BERT masih mampu mengenali keterkaitan makna yang cukup kuat. Temuan ini sejalan dengan Putri dan Ardiansyah (2023) yang menyatakan bahwa BERT memiliki kemampuan yang baik dalam memahami konteks Bahasa Indonesia [16], serta didukung oleh Holis et al. (2025) yang menjelaskan bahwa nilai cosine similarity yang mendekati 1 menunjukkan tingkat kesamaan makna yang sangat tinggi antar teks [17].

2. Performa Model DistilBERT

Model Model DistilBERT (paraphrase-multilingual-mpnet-base-v2) memperoleh rata-rata cosine similarity sebesar 0,5369 dan menempati posisi kedua setelah BERT, yang menunjukkan kemampuannya dalam menghasilkan representasi semantik yang cukup baik untuk mengukur kemiripan antar teks. Dengan standar deviasi sebesar 0,1319, DistilBERT menunjukkan tingkat konsistensi yang moderat serta mampu membedakan pasangan teks yang memiliki tingkat kemiripan berbeda dengan cukup baik. Selain itu, rentang skor yang diperoleh, yaitu dari 0,2167 hingga 0,7636, mengindikasikan sensitivitas model dalam mengenali variasi hubungan semantik antar teks. Temuan ini sejalan dengan Basbeth dan Fudholi (2024) yang menyatakan bahwa DistilBERT tetap mampu memberikan performa yang baik pada pemrosesan teks Bahasa Indonesia, sehingga dapat menjadi alternatif yang efektif ketika diperlukan keseimbangan antara kualitas representasi semantik dan efisiensi komputasi [18].

3. Performa Model SBERT

Model SBERT (paraphrase-multilingual-MiniLM-L12-v2) memperoleh rata-rata cosine similarity sebesar 0,5051, terendah di antara ketiga model yang diuji. Meskipun demikian, hasil ini perlu dipahami berdasarkan karakteristik arsitektur SBERT yang dirancang khusus untuk tugas semantic textual similarity melalui pendekatan Siamese Network dan contrastive learning, sehingga lebih diskriminatif dalam membedakan pasangan teks yang serupa dan berbeda. Hal ini ditunjukkan oleh standar deviasi tertinggi (0,1375) dan rentang skor terluas (0,2040-0,7419), yang mencerminkan sensitivitas model terhadap variasi kemiripan semantik. Performa yang lebih rendah diduga dipengaruhi oleh karakteristik data berupa teks pendek, informal, dan tidak baku, sedangkan model yang digunakan belum dioptimalkan untuk konteks Bahasa Indonesia tersebut. Temuan ini sejalan dengan Lubis et al. (2026) yang menyatakan bahwa efektivitas SBERT sangat bergantung pada kesesuaian domain dan karakteristik data, sehingga penggunaan model berbasis Bahasa Indonesia seperti IndoSBERT berpotensi memberikan performa yang lebih baik pada penelitian serupa [19].

KESIMPULAN

Hasil analisis kuantitatif menunjukkan bahwa minat remaja terhadap platform edukasi interaktif berbasis kuis berada pada kategori baik dengan rata-rata skor sebesar 3,86. Aspek kegunaan fungsional memperoleh nilai tertinggi dibandingkan aspek lainnya yang menunjukkan bahwa manfaat yang dirasakan pengguna menjadi faktor utama dalam meningkatkan minat penggunaan platform.

Berdasarkan perbandingan performa model NLP (Natural Language Processing) BERT menghasilkan rata-rata Cosine Similarity tertinggi sebesar 0,7666 yang diikuti model DistilBERT sebesar 0,5369 dan SBERT sebesar 0,5051. Hasil tersebut menunjukkan bahwa BERT memiliki kemampuan terbaik dalam memahami kemiripan semantik tanggapan responden pada penelitian ini.

Hasil regresi linear berganda menunjukkan nilai koefisien determinasi (R^2) sebesar 0,7516 yang berarti 75,16% variasi minat pengguna dapat dijelaskan oleh aspek tampilan visual, kemudahan operasional dan kegunaan fungsional. Hasil penelitian ini dapat dimanfaatkan sebagai dasar pengembangan platform dengan memprioritaskan peningkatan aspek kegunaan, kualitas fitur pembelajaran, serta pengalaman pengguna agar minat penggunaan dapat meningkat secara berkelanjutan. Secara keseluruhan kegunaan fungsional menjadi faktor utama minat pengguna, sedangkan BERT paling efektif mengidentifikasi kesamaan makna tanggapan. Temuan ini mendukung pengembangan website sesuai kebutuhan pengguna.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah mendukung penelitian ini. Semoga penelitian ini bermanfaat bagi pengembangan ilmu pengetahuan di masa mendatang.

References

ISSN 2714-7444 (online), <https://acopen.umsida.ac.id>, published by Universitas Muhammadiyah Sidoarjo

Copyright © Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

- [1] E. Rahayu Widyawati, "Pemanfaatan media pembelajaran berbasis teknologi sebagai alat pembelajaran kekinian bagi guru profesional IPS dalam penerapan pendidikan karakter menyongsong Era Society 5.0," *Proceedings Series on Social Sciences & Humanities*, vol. 10, p. 215, 2023, doi: 10.30595/pssh.v10i.667.
- [2] A. Izza and M. Munawir, "Hypertext based interactive media analysis in growing student reading interest through online learning," *Cyberspace: Jurnal Pendidikan Teknologi Informasi*, vol. 5, no. 1, p. 69, 2021, doi: 10.22373/cj.v5i1.8793.
- [3] I. K. Wardhani and I. Yuliati, "Efektifitas pendidikan kesehatan tentang dampak gadget pada kesehatan remaja di salah satu Paroki Surabaya," *JPK: Jurnal Penelitian Kesehatan*, vol. 11, no. 2, pp. 15–19, 2021, doi: 10.54040/jpk.v11i2.223.
- [4] G. Jenny and M. Belinda, "Dampak penggunaan gadget terhadap perkembangan perilaku anak remaja masa kini," *Sekolah Tinggi Teologi Excelsius Surabaya*, vol. 5, no. 1, pp. 7–10, 2021.
- [5] K. Hidayati, A. Rahmawati, and D. S. Wijayanto, "Effectiveness of using Google Sites-based learning media to improve critical and creative thinking abilities of vocational school students," *JPPI (Jurnal Penelitian Pendidikan Indonesia)*, vol. 10, no. 3, p. 632, 2024, doi: 10.29210/020243964.
- [6] H. Zhang and M. O. Shafiq, "Survey of transformers and towards ensemble learning using transformers for natural language processing," *Journal of Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-023-00842-0.
- [7] I. Karimah, M. N. Abdi, and M. Mufid, "Peran literasi digital, adaptabilitas dan self efficacy dalam memengaruhi kesiapan kerja Gen Z di era transformasi teknologi," *JMD: Jurnal Riset Manajemen & Bisnis Dewantara*, vol. 8, no. 1, pp. 41–52, 2025, doi: 10.26533/jmd.v8i1.1355.
- [8] R. J. Alexander and W. Suhendry, "Analisis pengaruh persepsi kemudahan penggunaan dan persepsi kegunaan terhadap minat penggunaan BCA Mobile Banking di Pontianak," *OBIS*, vol. 6, no. 2, pp. 18–27, 2024.
- [9] A. Iskandar, R. Fitriani, N. Ida, and P. H. S. Sitompul, *Dasar Metode Penelitian*. Yayasan Cendekiawan Inovasi Digital Indonesia, 2023.
- [10] E. Ismanto et al., *Revolusi Natural Language Processing: Ketika Mesin Memahami Bahasa*. PT Penerbit Qriset Indonesia, 2026.
- [11] J. F. U. Albab and A. N. Rohman, "Recommendation system Yogyakarta tourism using TF-IDF and cosine similarity methods with word normalizer," *Journal of Applied Informatics and Computing*, vol. 10, no. 1, pp. 935–945, 2026.
- [12] R. R. Putra, N. A. Putri, and A. D. Putra, *Teknik Cosine Similarity dan TF-IDF dalam Analisis Data*. Serasi Media Teknologi, 2024.
- [13] M. P. Syah, A. P. Wardani, and M. Idhom, "Perbandingan representasi teks TF-IDF dan BERT terhadap akurasi cosine similarity dalam penilaian otomatis jawaban berbasis teks," *Data Sciences Indonesia (DSI)*, vol. 5, no. 1, pp. 47–59, 2025.
- [14] M. Putri, T. E. Sutanto, and S. Inna, "Studi empiris model BERT dan DistilBERT analisis sentimen pada pemilihan presiden Indonesia," *The Indonesian Journal of Computer Science*, vol. 12, no. 5, pp. 2972–2980, 2023, doi: 10.33022/ijcs.v12i5.3445.
- [15] K. R. Sungkono, R. Sarno, M. C. Salsabila, and C. P. Dewi, "A graph-based method for merging business process models by considering semantic similarity," *International Journal of Intelligent Engineering & Systems*, vol. 16, no. 2, pp. 166–175, 2023.
- [16] N. A. R. Putri and A. Ardiansyah, "Analisis sentimen terhadap kemajuan kecerdasan buatan di Indonesia menggunakan BERT dan RoBERTa," *Jurnal Sains dan Informatika*, vol. 9, no. 2, pp. 136–145, 2023.
- [17] R. M. Holis, P. E. P. Utomo, and B. F. Hutabarat, "Semantic FAQ chatbot using SBERT (Sentence-BERT) and cosine similarity for academic services," *Brilliance: Research of Artificial Intelligence*, vol. 5, no. 2, pp. 915–922, 2025, doi: 10.47709/brilliance.v5i2.7027.
- [18] F. Basbeth and D. H. Fudholi, "Klasifikasi emosi data text bahasa Indonesia menggunakan algoritma BERT, RoBERTa, dan DistilBERT," *Jurnal Media Informatika Budidarma*, vol. 8, no. 2, pp. 1160–1170, 2024, doi: 10.30865/mib.v8i2.7472.
- [19] I. Lubis, H. Lubis, and I. N. Wahidah, "Implementation of semantic search in an academic repository using Sentence-BERT and FAISS," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 10, no. 2, pp. 1060–1069, 2026, doi: 10.33395/sinkron.v10i2.15940.